What do you think of my picture? Investigating factors of influence in profile images context perception

Filippo Mazza, Matthieu Perreira da Silva, Patrick Le Callet, Ingrid Heynderickx

What do you think of my picture?: Investigating factors of influence in profile images context perception

F. Mazza[a], M.P. Da Silva[b], P. Le Callet[b] and I.E.J. Heynderickx[c]

[a]LUNAM, Ecole Centrale de Nantes, IRCCyN UMR CNRS 6597, Nantes, FRANCE
[b]LUNAM, Université de Nantes, IRCCyN UMR CNRS 6597, Nantes, FRANCE
[c] Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, The Netherlands

ABSTRACT

Multimedia quality assessment has been an important research topic during the last decades. The original focus on artifact visibility has been extended during the years to aspects as image aesthetics, interestingness and memorability. More recently, Fedorovskaya proposed the concept of 'image psychology': this concept focuses on additional quality dimensions related to human content processing. While these additional dimensions are very valuable in understanding preferences, it is very hard to define, isolate and measure their effect on quality. In this paper we continue our research on face pictures investigating which image factors influence context perception. We collected perceived fit of a set of images to various content categories. These categories were selected based on current typologies in social networks. Logistic regression was adopted to model category fit based on images features. In this model we used both low level and high level features, the latter focusing on complex features related to image content. In order to extract these high level features, we relied on crowdsourcing, since computer vision algorithms are not yet sufficiently accurate for the features we needed. Our results underline the importance of some high level content features, e.g. the dress of the portrayed person and scene setting, in categorizing image.

Keywords: portrait images, content perception, high level features, social bias, social networks, crowdsourcing

1. INTRODUCTION

Social networks are filled with extremely different pictures, both in terms of objective quality and content. Some of them are high quality shots, but many of them are simpler consumer produced ones. The impact that these pictures can produce is also very different. Even low quality shots may have a huge impact, as the message they convey may be sometimes more important than their inherent quality. For example, family photo galleries or shoots from photo reporters can be invaluable, no matter their technical quality. The latter is probably even more the case with portraits, where different photos of the same person can deliver very different messages. The latter is often overlooked by people: our first impressions do not only influence online interactions, but also real life communications. For example job recruiters may look for online profiles to complement their opinions, as demonstrated by Manant et Al.[1] As stated by authors of,[2] it is important to consider image semantics where "human subjectivity" plays an important role. In this respect, current research related to image quality assessment seems to agree that considering only low level technical features is simply not enough, and so, we have to shift the attention towards high level features, more focused on "the domain of psychology". The latter concept is still broad and vague, but first attempts to outline it have been done by Fedorovskaya.[3] Following our previous work on portrait typologies,[4] in this paper we investigate which image features influence portraits' perception, where we more particularly assess the context to which the portrait is perceived to belong to. Portraits can suit different contexts and needs; for example, a portrait can fit well as a social network profile picture, but then may not fit a resume picture at all. Given previous considerations, we consider both low level and high level features, since we believe that the latter category greatly influences the evaluation of a portrait. Portrait images have been studied in different fields. Computer vision research focused on face features as input in machine learning algorithms to predict aesthetic assessment (as done by Li,[5] or more recently by Khan[6]). Research in cognitive

HVEI XX, 9 - 12 February 2015, San Francisco, California, USA
Author contact: filippo.mazza@irccyn.ec-nantes.fr

sciences tried to understand and modify image memorability.[7] Many of these approaches, however, adopted black box methods (i.e. ANN, SVM) and/or mathematical descriptors (i.e. HOG, GIST), losing interpretability of the results.[8] As our aim is to understand which interpretable features are influential, we preferred a white box approach. Hence, logistic regression is used to build a predictive model between portrait typology and image features, i.e., the final aim of our research.

2. METHODS

To evaluate which features influence portrait perception, we collected real online portraits related to three different contexts; namely to friends, work or dating purposes. We based this choice on current trends in social networks, nowadays focused on these three kinds of interactions (e.g., Facebook, Linkedin and Meetic). We then collected subjective assessments of the perceived context of each portrait. Successively we extracted both low and high level features from these portraits. The former ones were directly computed from pixel intensities (e.g., contrast), whereas the latter related to content interpretation were assessed subjectively. To have enough subjective evaluations of both the content category and the high level features, we opted for a large scale subjective campaign via crowdsourcing.

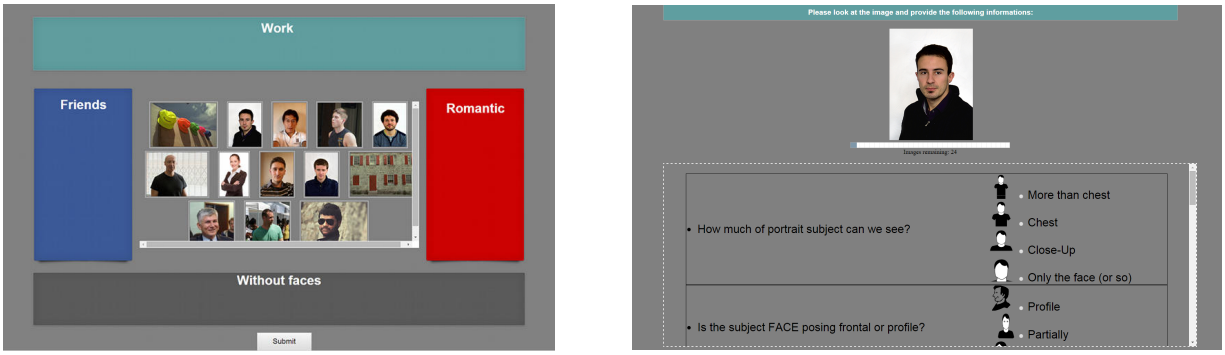
2.1 Portrait images

Many portrait databases exist online, but they are mainly dedicated to face detection-recognition,[9] or pose and expression estimation.[10] As such, they mainly focus on the face only (e.g., a neutral facial expression against a white background), or on full body portraits. In our research we were interested in face pictures, including a part of a person's torso and partly showing a background, as such portraits convey more complementary information regarding the context. Hence, we looked for amateur or semi-professional pictures, reflecting less formal and "posed" portraits - a characteristic that we suppose influence subjective context perception. For those reasons, we preferred to collect real online portraits from online networks and image sharing sites, collecting only publicly accessible images and taking care about licenses' restrictions *. We tried to select pictures related to the three chosen context categories: friends, work and dating purposes. While a huge number of portraits can be collected online, many problems arise, since many social networks do not allow - or strongly limit - pictures' disclosure. Facebook and Linkedin forbid using pictures without explicit permission of profile owners, and so, portraits of these databases weren't used in our experiment. Instead Twitter and Flickr offer APIs to access data. In the Flickr database we found portraits that we believe fit the categories friends and dating purposes. To gather work related portraits we adopted the website Elance.com, which is a work related social network for freelancers. In addition, we adopted 35 pictures from the LFW dataset,[9] suitable for the work related category. We avoided portraits of world famous people (e.g., world known politicians), since knowing their profession might influence the assessment of the fit to a category. A careful image selection was required. We manually checked content discarding all non portraits, non-adult subjects or inappropriate content. Extreme close-ups or too small pictures were also not selected, as content information was actually lacking. On the other hand, also pictures showing the context too clearly (e.g.,. showing signing a contract in a working environment) were discarded. To add more pictures to the ones we selected from the databases, we also used the best portraits we created for our previous experiment,[4] including both selfies and professional portraits. While it would have been possible to have a huge number of portraits, we limited the actual number for the experiment, since all pictures had to be assessed also on their high level features manually. Thus, as a compromise in terms of accuracy and reasonable time/cost for the experiment, we decided to use 216 collected pictures.

2.2 Crowdsourcing portrait evaluations

In order to gather a large amount of subjective evaluations we adopted crowdsourcing as in our previous work.[4] Crowdsourcing exploits the power of the web to outsource small tasks to people gathered online; it has already been investigated intensively and positively exploited in multimedia quality assessments. It is not the purpose of this work to dig into this technique; we refer the reader to our previous work, from which we adopted the same

*CC attribution licensed images have been taken, provided as found, citing sources. No information about depicted subject was given in order to preserve anonymity. Where no API was available and crawlers were forbidden, we gathered images manually



(a) (b)

Figure 1: Interfaces adopted for evaluating context (a) and high level features (b) of portraits.

platform and framework. In the current study, participants were asked to express for each portrait to which context it fits best. They were given three options, referring to friends, work and dating purposes, as shown in Figure 1. We then also asked participants to indicate the second best category of their choice. Implicitly, we then obtained their third choice too. This approach allowed us to have a ranking of the three contexts for each image. A fourth category named "without faces" was added for reliability check purposes as explained later. Classification was done through a web browser simply with "drag and drop", as was also explained in the provided instructions. Before starting the actual test we provided participants with a small anonymous questionnaire. We collected the participants' job, nationality, birth year, education level and if they were new to this kind of experiments. This information will be used in future work. At the end of the session we provided participants the possibility to leave a comment, in order to have useful feedback to monitor our experiment. It is well known that crowdsourcing participants are much less committed than participants in laboratory environments. Thus, particular attention on experiment duration and price paid is required, otherwise people may withdraw prematurely the test. Preliminary experiments showed that a good compromise to maximize efficiency while avoiding participants' withdrawal is providing 25 images per session. This number corresponds approximately to a test of ten minutes. With our previous experience in crowdsourcing and considering current prices paid for similar work, we decided to pay each participation 0.50$. Still participants can provide unreliable evaluations, both in good or bad faith (in the latter case, for example, to collect easy money). To have some control on that, we included two reliability checks - a.k.a. honeypots. In case of failure in any of the two honeypots participants were excluded from analysis and not paid. The first honeypot was a demographic question; the birth year was set to 1880 by default. Users that did not pay attention to the questions, then reported an impossible age. The second honeypot was part of the picture categorization process. Two images of the Toyama database,[11] not showing any person, were included in each session. The instructions clearly stated to mark such pictures as 'non faces', but many participants instead labelled them as portraits. While there might be multiple causes for such errors (e.g., misunderstanding the instructions, poor English comprehension or poor attention) we considered participants falling in one of the honeypots as outliers.

2.3 High level features evaluation

Different research addresses the analysis of features to describe image content. In particular, the authors of[12] investigate semantic features, mainly related to the setting of the scene (i.e., indoor/outdoor and which scene). Particular focus regarding features of humans is given in,[13] where visibility, clothing and appearance are considered. Regarding low level features and image quality assessment, a huge amount of research is reported in literature, and reviewed in.[6] The authors of[6] also adopted some of the proposed features to assess human portraits. Based on this literature we chose a subset of those features that we believed to be important in portrait context evaluation. We also added some high level features focused on the depicted context; in particular we included the depicted person's gender, the presence of glasses, his/her gaze direction and whether he/she was smiling in the picture. We didn't add emotion as a feature because we considered that smiling was a sufficient predictor to discriminate between emotions as in.[14] All features adopted in this research are summarized in table

LOW LEVEL	Aspect ratio, resolution, mean and standard deviation for hue, saturation and value, image contrast.
HIGH LEVEL	Face size [0-4], orientation [frontal,profile,half profile] and tilt [no,L,R]. Subject gender [M,F], dress [- ,informal,partially formal,formal], beard [0-5], glasses [no,normal, sunglasses], smile [0-2], eyes opened [0-2] and gaze [L,R,frontal]. Setting [indoor,outdoor,-], background type [- ,neutral,city,nature,room,office] and blurring [0-2]

Table 1: Adopted features, with possible values in brackets. Dash represents 'unknown value'.

1. Many of our features are categorical variables, as head tilt and orientation (left,center,right) or scene setting (don't know/indoor/outdoor), and some of these are ordinal as subject size (from small to big). Computer vision based high level features extraction can be time consuming and error prone. As first step, we prefer to focus on features importance and therefore, we adopted crowdsourcing to manually label them. The advantage of this approach was twofold: (1) it greatly speeded up the process and avoided errors, and (2) it offered a subjective opinion for some cognitive features that might be perceived differently between people (e.g., where a portrait has been shot). For this purpose, we followed the same crowdsourcing strategy, changing only the task for the participants. They now were provided with a web interface to evaluate each feature for each picture (as shown in Figure 1(b)). For some features (e.g., subject profile side) we gave guidelines in the form of icons near the answer options or the possibility to answer 'don't know' when in doubt. Some features were used jointly as honeypots to detect outliers - notably depicted person's gender and beard.

3. DATA ANALYSIS AND RESULTS

Thanks to crowdsourcing we quickly collected many subjective evaluations. More than 17000 subjective evaluations were collected from 440 participants, spread worldwide over 41 countries. More than 1000 people instead participated in the evaluation of the high level features. Despite these numbers, many participants withdrew the experiment before the end. We cannot say whether they considered their participation not worth the price or whether they experienced technical problems. Investigating experimental timings with server logs, we found that many users in remote regions, notably Eastern Asia, reloaded the test web page more than once in the first crowdsourcing campaign. This fact might indicate poor network connections, and as such, being unable to rapidly load images, even if scaled. However, most evaluations were correctly provided and useful for our analysis. Comments provided at the end of the test were dominantly positive; users noted that the experiment was clear and asked to be informed of future tests. Evaluations were checked for reliability using the honeypots. Around 20% of the participants failed to provide a reasonable birth year, and another 20% failed to properly indicate the non-portrait images. The corresponding submitted jobs were discarded and participants were notified. Finally, we obtained more than 8000 valid context rankings for further analysis. For the analysis of influential factors, we adopted a Logistic Regression, using our features as regressors and the context ranks as observations to fit. Each portrait picture was represented with its most often selected context ranking. Since the dependent variable had three possible discrete outcomes, we used a Multivariate Logistic Regression. The model, in case of a single observation can be written as:

$$y_n = \beta_0 + \sum_{k=1}^{K} x_{nk}\beta_k + \epsilon_n \tag{1}$$

where y is the dependent variable - the context rank for a particular category - x are our predictor values, β are the coefficients to be estimated and ϵ indicates the error term. Even if we expect our model to be more complex than linear, we adopted this method to have interpretable outputs for our features. We fitted three independent models, one for each context, adopting the ranks of that context for all portrait pictures as responses. We used a majority strategy to define the value of the high level features for each portrait picture. So, if a picture was reported to be shot outdoor by 90% of the participants, it was considered as such for the labelling of the high level features. While this subjectivity may be valuable for our study at this stage, we also consider adopting the raw data for the high level features in future work. All features were normalized (especially the low level ones) to the same mean and range, so that the computed coefficients could reflect actual relevance weights. The results of

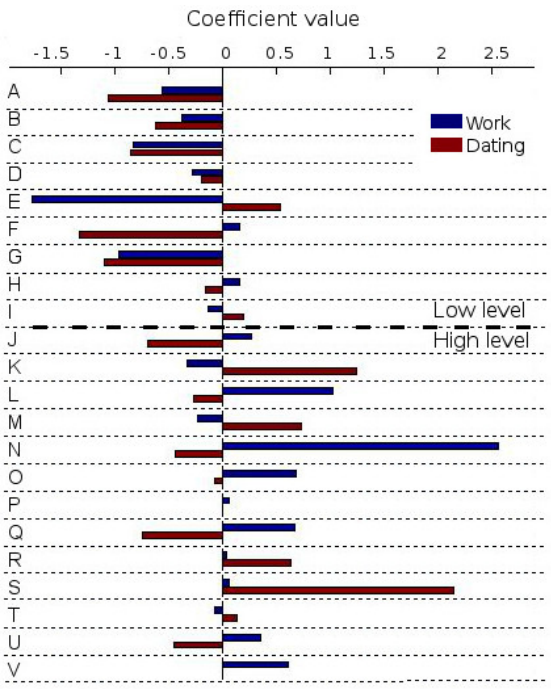


Figure 2: Features regression coefficients for first choice. Friend context has been taken as reference.

Feature	Work	Dating
A: Aspect ratio		
B: Resolution		
C: Hue μ		
D: Hue σ		
E: Saturation μ	*	
F: Saturation σ		
G: Value μ		
H: Value σ		
I: Contrast		
J: Face size		
K: Face orientation		**
L: Indoor/outdoor	**	
M: Beard		
N: Dress	***	
O: Glasses		
P: Smile		
Q: Sight		
R: Background		
S: Gender		***
T: Head tilt		
U: Background blur		
V: Eyes opened		

Table 2: Features significance. *=$p < 0.05$, **=$p < 0.01$, ***=$p < 10^{-3}$

the logistic regression fit are shown in Figure 2. The Friends context was selected as reference in our model. This means that each computed coefficient expresses the influence of a feature on the relative chance that a portrait picture is perceived in another context (i.e., Work or Dating) than the reference, where this chance is expressed in log odds. So for example, an increment of one unit in the feature "Dress" increases β_k times the relative log odds of a portrait picture being perceived as Work context, where β_k is the coefficient related to the feature "Dress" in the model of the Work context - assuming everything else being equal. The adopted regression method also provides a measure of the statistical influence of each feature in the model via their p-value. These p-values are shown in table 2. They illustrate that the prediction model for a portrait picture having a Work context is significantly affected by the dress (N) that the person in the picture wears, as well as by the portrait setting (L) and by the low level feature of mean saturation (E) in the picture. As expected, a formal dress increases the appropriateness of the portrait for working purposes, while instead an increase in saturation decreases it. For predicting the appropriateness of a portrait for dating purposes instead the gender of the portrayed person, and his/her face orientation are statistically significant contributors. In particular portraits depicting female persons are reported as more appropriate for dating purposes. In this respect, it is interesting to check if the gender of the participant generates a bias to the influence of the gender of the portrayed person. To this extent we adopted permutation test as in our previous work.[4] The factor investigated with a chi-square test was the gender of the participant, as obtained from the demographic questionnaire , while the dependent variable was the context evaluation. Note that this analysis considered all the raw evaluations before the majority strategy. Permutation tests underlined that an effect is present ($p < 0.05$). To analyse if the effect is present especially on one context evaluation, we repeated permutation test considering one context (i.e. work) against the other two (i.e. friends+dating). Gender was found statistically influential only for the appropriateness of portraits for dating purposes, underlining that portraits were evaluated differently between men and women only in the context of dating. This finding pointed out that gender should be more explicitly considered in future research, for example, by only considering evaluations, for which the gender of the portrayed person and the gender of the participant are the same.

4. DISCUSSION

In this work we addressed the problem of determining influential factors for perceived context of portrait pictures. Our analysis considered some classical low level features as well as higher level features related to image interpretation. It is clear that considering all possible factors of influence in only one analysis is a challenging task. However, isolating them to determine each feature's separate impact is at least as hard and time consuming. We approached the problem considering features we supposed mostly important and we showed how multiple features can be addressed simultaneously. To this extent we gathered more data for a statistical analysis, and so, adopted crowdsourcing not only to determine the appropriateness of a portrait picture for a given context, but also to evaluate the high level features. In fact, extracting these only with computer vision tools can be error prone. We used a white box approach for the statistical analysis that even if less accurate than a black box approach, yet allows easier interpretation of the results. While the proposed analysis is not able to explain all variability in the subjective assessments, some statistically relevant features have been underlined and greatly helped to discriminate between portrait contexts. Our first results are consistent with expectations from empirical experience. In particular the dress of the portrayed person is shown to be discriminative for the likelihood of a portrait to be perceived as work related. Also the gender of the portrayed person appeared to be influential, more particularly for the likelihood of a portrait to be perceived as dating related. The latter, however, was also dependent on the gender of the participant; this element should be taken into account in future works. Interestingly the background interpretation was not as influential as we expected. Instead background interpretation only contributed very little to our model. Future research will deal with online portraits retrieval to adopt a broader scale analysis. With these results in mind, we will focus on computer vision efforts on features underlined as influential.

REFERENCES

[1] Manant, M., Pajak, S., and Soulié, N., "Do recruiters 'like' it? Online social networks and privacy in hiring: a pseudo-randomized experiment," Tech. Rep. 56845, Munich Personal RePEc Archive (MPRA) (2014).

[2] Joshi, D. and Datta, R., "Aesthetics and emotions in images," *IEEE Signal Process. Mag.* **28-5**(SEPTEMBER 2011), 94–115 (2011).

[3] Fedorovskaya, E. A. and De Ridder, H., "Subjective matters: from image quality to image psychology," in [*Hum. Vis. Electron. Imaging XIII*], **8651**, 86510O–86510O–11 (2013).

[4] Mazza, F., Perreira Da Silva, M., Le Callet, P., and Da Silva, M. P., "Would you hire me? Selfie portrait images perception in a recruitment context," in [*Hum. Vis. Electron. Imaging XIX*], **9014**, 90140X (2014).

[5] Li, C., Gallagher, A., Loui, A. C., and Chen, T., "Aesthetic quality assessment of consumer photos with faces," in [*2010 IEEE Int. Conf. Image Process.*], 3221–3224, IEEE (Sept. 2010).

[6] Khan, S. and Vogel, D., "Evaluating visual aesthetics in photographic portraiture," *Proc. Eighth Annu. Symp. Comput. Aesthet. Graph. Vis. Imaging (CAe '12)* , 55–62 (2012).

[7] Khosla, A., Bainbridge, W. a., Torralba, A., and Oliva, A., "Modifying the Memorability of Face Photographs," in [*Int. Conf. Comput. Vis.*], (2013).

[8] Marchesotti, L., Murray, N., and Perronnin, F., "Discovering Beautiful Attributes for Aesthetic Image Analysis," *Int. J. Comput. Vis.* (November) (2014).

[9] Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E., "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," tech. rep., University of Massachusetts (2007).

[10] Sim, T., Baker, S., and Bsat, M., "The CMU Pose, Illumination, and Expression (PIE) database," in [*Proc. Fifth IEEE Int. Conf. Autom. Face Gesture Recognit.*], (1), 53–58, IEEE (2002).

[11] University of Toyama, "Toyama MICT Image Quality Evaluation Database," (2010).

[12] Totti, L., Costa, F., Avila, S., Valle, E., Meira, W. J., and Almeida, V., "The Impact of Visual Attributes on Online Image Diffusion," in [*ACM Web Sci. Conf.*], (2014).

[13] Isola, P., Xiao, J., Torralba, A., and Oliva, A., "What makes an image memorable?," *Cvpr 2011* , 145–152 (June 2011).

[14] Xue, S.-f., Tang, H., Tretter, D., Lin, Q., and Allebach, J., "Feature design for aesthetic inference on photos with faces," in [*2013 IEEE Int. Conf. Image Process.*], 2689–2693, IEEE (Sept. 2013).